DOCUMENT RESUME

ED 204 354 TM 810 295

AUTHOR Schwandt, Thomas A.
TITLE Defining Rigor and Relevance in Vocational Education Evaluation.
PUB DATE 17 Apr 81
NOTE 28p.: Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Definitions: *Epistemology: Ethnography: *Evaluation Methods: *Program Evaluation: Research Design: Vocational Education
IDENTIFIERS *Relevance (Evaluation): *Rigor (Evaluation)

ABSTRACT

The terms "rigor" and "relevance" most often surface in discussions of methodological adequacy. Assessing epistemological relevance is equivalent to answering the question, "Is this particular research question worth asking at all?" Epistemological rigor refers to the properties of a "researchable" problem. If one accepts the proposition that different kinds of questions require different kinds of methodologies, then the question of methodological relevance can be asked as, "Is this particular method (or model) appropriate to the questions that I am trying to answer?" Methodological rigor is a determination of whether the method selected meets certain standards for a "good" or "trustworthy" method, a "sound" design, or an "appropriate" type of data analysis. It is frequently argued that one must choose between rigorous and relevant methods. This argument demonstrates a confounding of the notions of methodological rigor and methodological relevance. Rigor and relevance are not necessarily inversely related. These issues are being discussed by vocational evaluators, and warrant additional investigation. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED204354

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

Defining Rigor and Relevance in Vocational Education Evaluation

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T. A. Schwandt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Thomas A. Schwandt
Department of Vocational Education
Smith Research Center
Indiana University
Bloomington, Indiana 47405

Paper presented at the 1981 annual meeting of the
American Educational Research Association
Los Angeles, California, April 17, 1981

TM 810 295

DEFINING RIGOR AND RELEVANCE IN VOCATIONAL EDUCATION EVALUATION

The terms 'rigor' and 'relevance' most often surface in discussions of methodological adequacy. If they have a 'classical' meaning, then, most likely, the phrase 'rigor versus relevance' refers to the tradeoffs involved in designing an experiment that has both high internal validity (rigor) and high external validity (relevance) (Campbell and Stanley, 1966). Within the context of the current methodological debate, these two terms have been used in a rather general way to characterize the differences between 'hard' data and traditional, scientific, or quantitative methodology which is rigorous and 'soft' data and less conventional, naturalistic, or qualitative methodology which is relevant.

I initially intended to investigate vocational education evaluation models, methods, and frameworks in view of their treatment of these two dimensions of methodological adequacy. Yet, attempts to clarify the meaning of the terms 'rigor' and 'relevance' revealed that they have an epistemological as well as a methodological meaning.

Hence, in what follows, I propose a more expanded analysis of rigor and relevance. Four distinct, though interrelated, notions of rigor and relevance are identified: epistemological relevance, epistemological rigor, methodological relevance, and methodological rigor. Each is explained and an attempt is made to illustrate the treatment of each in the literature on vocational education evaluation. The paper concludes with a discussion of the need to further investigate each dimension.

Defining Rigor and Relevance

Logical and conceptual analysis of the notions of rigor and relevance in the context of social science research and evaluation reveals an epistemological and a methodological usage of each term. These four aspects of rigor and relevance are explained below.

Epistemological Relevance

It is commonplace to characterize the major criteria for adequate research problems as relevance and fruitfulness. For example, Steiner (1978) explains that for an educational problem to be relevant it must be generative of scientific, philosophical, or praxiological knowledge about education. The mark of a fruitful problem is that it must be capable of leading to the extension of knowledge. Assessing epistemological relevance is thus equivalent to answering the question, "Is this particular research question worth asking at all?"

We might extend this notion of epistemological relevance in research to the domain of evaluation in the following way. To determine whether a particular evaluation question is worth asking (or whether a particular evaluation is worth pursuing), we must first answer two questions: (1) What is to be evaluated?, and (2) Why is it to be evaluated? These questions must be answered to the satisfaction of stakeholders in any given evaluation before questions of how to proceed are proposed. Failure to adequately specify the evaluand--the entity being evaluated (Scriven, 1979)--and to identify the intended uses and users of evaluative information will likely result in an inaccurate evaluation of little use to anyone.

Determining epistemological relevance need not be narrowly conceived as a task unique to positivistic science. The following two approaches to assessing epistemological relevance originate in quite different points of view regarding the nature of evaluation. The first approach, proposed by Joseph Wholey and his colleagues at the Urban Institute (Wholey, 1975, 1977; Horst, et al., 1974) is compatible with the traditional view of evaluation as 'evaluation research'. The approach, known as "evaluability assessment," requires clarifying and defining the evaluand from the perspectives of both the user and the evaluator. The major elements of an evaluability assessment are characterized by Wholey (1977) as:

1. Determining the boundaries of the problem/program, i.e., what is it that is to be analyzed?, what are the program objectives?

2. Gathering information that defines program objectives, activities, and underlying assumptions.

3. Developing a model of program activities and objectives from the point of view of the intended user of the evaluation information.

4. Determining to what extent the definition of the program, as contained in the model, is sufficiently unambiguous to permit a useful evaluation.

5. Presentation of the above information to the intended user of the evaluation and determination of next steps to be taken.

Though quite antithetical to the general approach of evaluation research, Guba's (1978) commentary on the methodology of naturalistic inquiry in evaluation also addresses the question of epistemological relevance. In surfacing the concerns and issues of relevant parties to the evaluation, the naturalistic investigator is engaged in a process of determining what is to be evaluated and why it is to be evaluated. Having cycled through repeated

phases of discovery and verification, the evaluator possesses a preliminary set of categories of information. Guba suggests that considerations of salience, credibility, uniqueness, heuristic value, feasibility, and materiality be employed to prioritize the categories and thereby focus the inquiry. As a result of this process, the evaluator is able to pursue those categories of concerns and issues "most worthy of further exploration."

It should be apparent from these two examples that, regardless of one's philosophical orientation regarding the nature of evaluation, assessment of epistemological relevance is a critical first step in evaluation. Though the techniques of the evaluation researcher and the naturalistic investigator are quite different, both aim at clarifying the nature of the evaluand and surfacing the concerns of potential users of evaluation information.

Epistemological Rigor

When we discuss the properties of a 'researchable' problem, we are speaking of epistemological rigor. As was the case with epistemological relevance, this notion of rigor is important regardless of the particular philosophical orientation of the researcher or evaluator. However, 'scientific' and 'naturalistic' inquirers assess the dimension of epistemological rigor in quite different ways.

Where evaluation is viewed as an extension of scientific research, the assessment of epistemological rigor is quite straightforward. Here, rigor refers to the extent to which evaluation questions are cast in a form that is measurable or testable. Assessment of rigor is largely context-independent and a priori. It involves adequately specifying the empirical referents for

and the connections between variables contained in the statement of the research or evaluation problem. Darcy (1979, 1980) illustrates this process in the context of vocational education outcomes evaluation. Beginning with a meaningful outcome statement, he explains that (1) this statement must be translated into a hypothesis with careful specification of dependent and independent variables, (2) the entity on which the outcome is to be observed must be delineated, and (3) empirical indicators of the outcome must be specified. Outcome statements so formulated meet the test of epistemological rigor.

Where evaluation is placed more in the tradition of ethnography (e.g., Guba, 1978; Patton, 1980; Filstead, 1979) the determination of epistemological rigor is no less important, yet the process is less apparent because problems are not so carefully circumscribed prior to the investigation. Here, the assessment of epistemological rigor is largely context-dependent and a posteriori. That is, epistemological rigor is not assessed at the outset of an investigation by determining whether or not a problem is researchable, rather rigor is assessed near the conclusion of the investigation by determining whether or not a problem has been adequately researched (investigated).

Naturalistic or ethnographic evaluation approaches begin with problems that are not largely delimited. Hence, the naturalistic evaluator defines epistemological rigor in terms of whether the limits to an investigation of a problem have been reached. Guba (1978) describes this process as one of reaching "closure" by applying the criteria of "exhaustion of resources," "saturation," "emergence of regularities," and "overextension" to the activity

of collecting information. If these signals for closure appear, then the naturalistic evaluator is reasonably certain that epistemological rigor has been attained.

Both approaches focus on setting the boundaries that define a researchable problem. In the case of evaluation research, these boundaries are defined in terms of conditions for stating the problem. In naturalistic evaluation, these boundaries are defined in terms of outer limits that signal completion of an investigation.

Methodological Relevance

If we accept the proposition that different kinds of evaluation questions require different kinds of methodologies, then the question of methodological relevance can be asked as, "Is this particular method (or model) appropriate to the questions that I am trying to answer?" Assessing methodological relevance is largely a matter of determining the tradeoffs, in terms of strengths and weaknesses, of methods that are available to the evaluator.

Questions of methodological relevance are largely means-end questions. It is only after knowing what we are trying to discover that we can decide how to proceed. For example, if we wish to test causal hypotheses, then we might choose an experimental design. If we wish to act as the surogate eyes and ears of decision makers who desire information about what really takes place in a program, then we might choose a case study approach.

Determining methodological relevance requires (1) a review of the conditions which must be present to facilitate the use of a given method and (2) a careful consideration of the intrinsic strengths and weaknesses

of the method. For example, in order to make use of an experimental or quasi-experimental evaluation design, conditions such as the following must obtain: a clear, precise statement of intended program results; a reasonably 'controlled' program setting; a reasonably 'uniform' treatment across participants and over time; a large enough sample; and an ability to select and assign individuals randomly to treatment and control groups, or the availability of a comparison group. Likewise, the measurement of program effects by means of objective, standardized instruments such as aptitude tests, achievement tests, or attitude scales requires that there be a program logic exhibiting valid linkages between the program's goals, the treatment delivered, and the instruments used to measure outcomes.

To understand the second objective--the process of weighing the intrinsic merits of a given technique--consider the following review of the technique of documentary analysis. This method involves the analysis of written program materials--e.g., interim reports, internal memoranda, activity logs, etc.--to gain a clearer insight of program planning and operation. Its strengths are that it is entirely unobtrusive and nonreactive, that the documents themselves are unchanging and express the perspectives of their authors in the authors' own natural language. On the other hand, among its weaknesses are that the document may not be representative, it is usually uniperspectival, represents unique events, and may be temporally and spatially specific. In a similar way, every method available to the evaluator can be scrutinized for its intrinsic adequacy or merit.

The activity of determining methodological relevance is clearly not a simple process. The choice of one method over another involves the evaluator

and other parties to the evaluation in a series of trade-offs for which no set of rules will suffice. It may be tempting to say that the determination of relevance should be based on the principle of maximizing utility. But that raises the question of how we are to measure utility and for whom. A more plausible approach may be to argue that instead of seeking optimal or maximal solutions to the problem of methodological relevance, we should adopt a strategy of "satisficing" (Simon, 1976)--choosing methods which are not necessarily the best but 'good enough' given the goals of the evaluation, the limitations of the methods themselves, the problems inherent in the particular evaluation situation, and the needs of relevant parties to the evaluation.

Questions of methodological relevance naturally raise the possibility of combining methods in a single study. The rationale for the use of multiple methods is captured nicely by Webb, et al. (1966):

> Once a proposition has been confirmed by two or more measurement processes, the uncertainty of its interpretation is greatly reduced. The most persuasive evidence comes through a triangulation of measurement processes. (p. 3, emphasis added)

Denzin (1978) further suggests that there are four types of triangulation available: (1) data triangulation--using a variety of data sources, (2) investigator triangulation--using several different evaluators, (3) methodological triangulation--using several different methods to examine the same questions so that the flaws of one method can be compensated for by the strengths of other methods, and (4) theory triangulation--using multiple perspectives to interpret

the same set of objectives. However, though the rationale may be well-established, actually effecting a methodological mix in a given study is a complicated matter (cf. Trend, 1979) requiring a great deal more investigation.

Methodological Rigor

Unlike methodological relevance which is an assessment of instrumental worth, methodological rigor is a determination of intrinsic merit. When we ask, "Is this method/evaluation design/type of data analysis rigorous?" we are asking whether it meets certain agreed upon standards for a 'good' or 'trustworthy' method, a 'sound' design, or an 'appropriate' type of data analysis.

Until very recently, the only available and agreed upon standards for methodological rigor were the canons for what constituted rigorous scientific inquiry. For example, in their review of federal evaluation studies, Bernstein and Freeman (1975) developed a composite index of scientific standards for measuring the quality (rigor) of evaluation research. Their rating scheme is shown below in Table 1.

Insert Table 1 about here

The Bernstein and Freeman index is fairly representative of the types of standards currently in use for judging the methodological rigor of both research and evaluation studies. Similar sets of standards are commonly employed in assessing the internal and external validity of research designs (Campbell and Stanley, 1966; Cook and Campbell, 1979), the psychometric properties of measurement devices (Guilford, 1954; Nunnally, 1978), etc.

Standards for assessing the methodological rigor of evaluation studies conducted in a naturalistic or ethnographic mode are far less well developed. A recent paper by Guba (In press) represents one of the first attempts to specify standards for naturalistic studies. Table 2 below displays the criteria which Guba proposes for assessing the methodological rigor ("trustworthiness") of naturalistic inquiries. Guba defines the naturalistic investigator's analog for criteria such as objectivity ("confirmability"), reliability ("dependability"), generalizability ("transferability"), and internal validity ("credibility") and lists and briefly explains methods that might be used to determine whether these criteria have been met.

Insert Table 2 about here

Efforts such as this to specify the standards for judging not only the design but the product of naturalistic inquiries are indispensable in view of the growing interest in the use of naturalistic and ethnographic methods.

Relationships Between Rigor and Relevance

The preceding four categories of rigor and relevance--epistemological rigor, epistemological relevance, methodological relevance, and methodological rigor--have been presented in their most logical sequence. It should be apparent that efforts to frame a question in a rigorous way should commence only after it is determined what it is that we are asking. Likewise (assuming the existence of standards for judging the rigor of both quantitative and qualitative methods), it is reasonable to believe that questions of which

method to use can be settled before examining how these methods might be used (or whether they have been used) in the most rigorous fashion.

This analysis of rigor and relevance has also demonstrated that these notions are not necessarily inversely related. In other words, an increase in relevance need not result in a decrease in rigor and vice-versa. To be sure, the four dimensions of rigor and relevance are not orthogonal. For example, an assessment of epistemological relevance informs the assessment of methodological relevance. Nevertheless, one need not always tradeoff rigor for relevance.

Tradeoffs between rigor and relevance frequently (and quite inappropriately) characterize the choice of evaluation and research methods. It is argued that one must choose between rigorous and relevant methods. This demonstrates a confounding of the notions of methodological rigor and methodological relevance. As was discussed above, the relevance of any method can be assessed with respect to the goals of the research or evaluation. Methods are instrumentalities; the suitability of a method for meeting the goals of inquiry determine its relevance. Rigor is another matter. Once a relevant method has been chosen, steps can be taken to ensure the rigorous use of that method. The only well-developed and agreed upon standards for rigor apply to the use of quantitative methods. We have only recently begun to investigate standards for rigor that govern the use of qualitative methods. It is not the case that qualitative methods are inherently non-rigorous (and hence, somehow more relevant), but that, at present, we are uncertain of how to judge whether they have been used in a rigorous fashion.

Rigor and Relevance in Vocational Education Evaluation

Given the diversity of approaches to and methods for evaluating vocational education programs, it is hazardous to offer general statements regarding the extent to which the enterprise of vocational education evaluation is addressing these questions of rigor and relevance. However, it is commonplace to find questions of rigor and relevance addressed to varying degrees within the context of particular methods or approaches. From these discussions there emerge several central tendencies which are discussed below.

Epistemological relevance is emerging as a primary concern after several years of evaluation efforts. For example, following a two-year study of vocational education outcomes by the National Center for Research in Vocational Education (Darcy, 1979, 1980) it was recommended that:

> In planning evaluation studies, care should be taken to determine clearly what is to be evaluated and what criteria, data, and evaluation standards are to be used. (Darcy, 1980, p. 70)

This recommendation stems from several findings of this study which point to shortcomings in assessing epistemological relevance: (1) Terms such as 'outcomes,' 'outcome measures,' 'program goals,' and 'program benefits' lack precise definition, (2) It is not clear what is being evaluated--outcomes, groups of students, programs, etc., (3) There is little appreciation of the range, diversity, and complexity of possible outcomes, and (4) The relative importance of outcomes vis-a-vis other types of evaluation has not been well addressed.

The issue of epistemological relevance has been alluded to in other ways as well. In discussing the collection of evaluative data by means of a standardized vocational education data system, Drewes (1978) noted that answering questions of why data are to be collected (a dimension of assessing epistemological relevance) will determine the use and utility of such a system. Kievit (1978) sought to lay out a rationale for linking kinds of evaluative data to the values perspectives of potential users, thereby addressing the question of why vocational education programs are in need of evaluation.

Determining epistemological rigor has always been, and will likely remain, a major concern of vocational evaluators. For example, Lecht's (1974) discussion of indicators of vocational program success can be viewed largely as an attempt to address questions of epistemological rigor in the definition and measurement of those indicators. Most recently, the link between epistemological rigor and relevance has been demonstrated in the vocational education outcomes study noted earlier. The study attempted to document epistemological relevance for outcomes evaluation by requiring (1) a clear rationale for the choice of an outcome, (2) evidence of the appropriateness of a given outcome as a basis for program evaluation, (3) illustration of the potential impact of results, and (4) identification of relevant audiences for evaluative information. As noted earlier, the study then addressed epistemological rigor by indicating how outcome statements are to be translated into empirical measures of outcomes. Owing to the relatively recent importation of qualitative techniques to vocational

evaluation, there are no commentaries on procedures for establishing epistemological rigor in ethnographic or naturalistic vocational evaluation studies.

However, as alternative methodologies for evaluating social programs have found their way into evaluations of vocational programs, discussions of methodological relevance have emerged. Bolland (1979), for example, briefly addresses the question of methodological relevance by listing the relative strengths and weaknesses of various data gathering techniques in her review of vocational education outcomes studies. Grasso (1979) discusses the suitability of impact evaluation for meeting the evaluation requirements spelled out in the 1976 vocational education legislation. Spirer (1980) points to the utility of the case study method in vocational education evaluation. Bonnet (1979) discusses alternative methods for measuring the outcomes of career education in view of the outcome goals set by the Office of Career Education. Finally, Riffel (1980) recently offered a very reflexive presentation on the utility of the case study approach in the Vocational Education Study, and Pearsol (1980) commented on the implications of combining quantitative and qualitative methods.

Methodological rigor has perhaps been the most frequently addressed aspect of rigor and relevance in vocational education evaluation. Most, if not all, of these discussions are concerned with specifying standards for scientific rigor as it is commonly perceived in the research community. Hence, Bolland (1979) specifies eight basic components of a sound research report. Morell (1979) and Franchak and Spirer (1978) address design and

statistical issues in the application of follow-up research to vocational evaluation. Borgatta (1979) discusses the requirements for 'good' experimental and quasi-experimental designs.

Several papers address questions of methodological rigor and relevance simultaneously in reviewing a particular research technique. Pucel (1979) addresses questions of methodological relevance by pointing to the types of questions that can be answered through longitudinal studies. He also focuses on aspects of methodological rigor in the use of the method (e.g., solving problems in implementation, specifying types of data to be collected, etc.). Likewise, Franchak, et al. (1980) seek to demonstrate methodological relevance by linking the use of longitudinal methods to critical data needs in vocational education, and they address problems of rigor in reviewing basic strategies and procedures for longitudinal studies. Similarly, several publications in the Career Education Measurement Series (e.g., McCaslin, et al., 1979; McCaslin and Walker, 1979) address both methodological rigor and relevance in discussing the selection, evaluation, and design of instruments to evaluate career education.

In summary, the importance of addressing the issues of epistemological and methodological rigor and relevance can be seen in Lee's (1979) discussion of the factors governing the use of evaluation data. Lee identifies the following five factors: (1) availability (making evaluation data available to users in a way that can be readily understood), (2) reliability, (3) credibility, (4) utility (collecting, analyzing, and interpreting evaluation data in view of their potential uses), and (5) consistency (collecting,

analyzing, and making available data within the boundaries of possible action). It is possible to recast these factors as functions of addressing rigor and relevance in evaluation studies. Thus, failure to address questions of epistemological relevance may lead to problems in consistency and utility; failure to address questions of epistemological rigor and questions of methodological rigor may result in problems with reliability and credibility; finally, failure to address questions of methodological relevance may lead to problems with utility and availability.

Avenues for Future Study

All four dimensions of rigor and relevance discussed in this paper warrant further attention by the community of vocational education evaluators and researchers. Epistemological relevance--determining what is to be evaluated and why--must clearly be our foremost concern. Premature focus on the selection of appropriate methods will likely encourage the approach of 'solutions in search of problems.' That is, we may attempt to fit existing (and new and developing) evaluation strategies to particular vocational education evaluation problems without first understanding what it is we wish to know and why. There should be no equivocating: arguing that we have the 'right' solution but the 'wrong' problem is simply an argument for the wrong solution. We should not hesitate to retreat from solutions to make a more careful diagnosis of the problem.

Attending to epistemological rigor presents us with two different types of problems. It appears that we are fully in possession of the knowledge of

what constitutes a testable or measurable problem from a positivistic perspective. What we need are more attempts, such as that demonstrated in the vocational education outcomes study, to apply this knowledge to particular evaluation questions. On the other hand, as naturalistic inquiry becomes increasingly relevant to evaluations of vocational education programs, we will need to devote our efforts to specifying procedures for determining the boundaries of such investigations. The lack of a priori constraints, characteristic of this approach, does not imply a total lack or regard for constraints which demonstrate rigor in the investigation of problems.

Methodological relevance--including both an assessment of the intrinsic merits of methods and an investigation of the possibilities for combining methods--demands our most careful attention, lest the choice of methods becomes simply a matter of what is currently in vogue. We must guard against the normative appeal of certain established methods as being the most (or the only) 'rational' strategies and investigate the contextual limits governing the scope of these strategies. We must be careful not to mistake evidence of the inapplicability of certain methods as simply problems with implementation.

Finally, in the area of methodological rigor, we lack little knowledge of traditional standards for assessing the scientific adequacy of quantitative methods and experimental designs. Yet, we are largely ignorant of how to judge the merit of case studies, emergent designs, and similar methods and tools associated with naturalistic or ethnographic inquiries.

In general, we need to become more open and public about our discussions or rigor and relevance. There are relatively few accounts of the conduct of

vocational education inquiries that are reflexive. Reflexivity refers to the capacity of thought to bend back upon itself, to become an object to itself (Ruby, 1980). To be reflexive is not the same as being self-conscious or reflective. Most evaluators and researchers are probably self-conscious, yet that kind of awareness remains private knowledge for the inquirer, detached from the product of his or her inquiry. There are relatively few accounts of inquiry in which inquirers reveal the epistemological and axiological assumptions which caused them to choose a particular set of questions to investigate, to seek answers to those questions in a particular way, and, finally, to present their findings in a particular way. By engaging in this kind of reflexivity about our research and evaluation, we are more likely to address critical issues in rigor and relevance.

TABLE 1

Criteria for Assessing the Quality

of Evaluation Research*

Criterion	Rating
Sampling	1 - systematic 0 - nonrandom, cluster, or nonsystematic
Data Analysis	2 - quantitative 1 - qualitative and quantitative 0 - qualitative
Statistical Procedures	4 - multivariate 3 - descriptive 2 - ratings from qualitative data 1 - narrative data only 0 - no systematic material
Design	3 - experimental or quasi-experimental with randomization and control groups 2 - experimental or quasi-experimental without both randomization and control groups 1 - longitudinal or cross-sectional without control or comparison 0 - descriptive, narrative
Sampling	2 - representative 1 - possibly representative 0 - haphazard
Measurement Procedures	1 - judged adequate in face validity 0 - judged less than adequate in face validity

*Bernstein, Freeman, 1974, pp. 100-101

TABLE 2

Criteria for Assessing the Trustworthiness
of Naturalistic Inquiries*

Aspect of Method, Study, Procedure	Naturalistic Term	Methods for Determining Whether Criteria Are Met
Truth Value	Credibility	Prolonged engagement at site, Peer debriefing, Triangulation, Member checks, Collection of referential adequacy materials
Applicability	Transferability	Theoretical/purposive sampling, Collection of "thick" descriptive data
Consistency	Dependability	Overlap methods, Stepwise replication, Establish "audit" trail
Neutrality	Confirmability	Triangulation, Confirmability audit

*Guba, In press, passim

References

Bernstein, I.N., & Freeman, H.E. Academic and Entrepreneurial Research. New York: The Russell Sage Foundation, 1975.

Bolland, K.A. Vocational Education Outcomes: An Evaluative Bibliography of Empirical Studies. Columbus, OH: The National Center for Research in Vocational Education, 1979.

Bonnet, D.G. Measuring Student Learning in Career Education. In T. Abramson, C.K. Tittle, L. Cohen (Eds.), Handbook of Vocational Education Evaluation. Beverly Hills, CA: Sage, 1979.

Borgatta, E.F. Methodological Considerations: Experimental and Non-experimental Designs and Causal Inference. In T. Abramson, et al., (Eds.), Handbook of Vocational Education Evaluation. Beverly Hills, CA: Sage, 1979.

Campbell, D.T., & Stanley, J.C. Experimental and Quasi-experimental Designs for Research. Chicago: Rand McNally, 1966.

Cook, T.D., & Campbell, D.T. Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally, 1979.

References, continued

Darcy, R.L. Some Key Outcomes of Vocational Education. Columbus, OH: The National Center for Research in Vocational Education, 1980.

Darcy, R.L. Vocational Education Outcomes: Perspective for Evaluation. Columbus, OH: The National Center for Research in Vocational Education, 1979.

Denzin, N.K. The Research Act. New York: McGraw Hill, 1978.

Drewes, D.W. Outcome Standardization for Compliance or Direction: The Critical Distinction. Paper presented at the National Conference on Outcome Measures, Louisville, Kentucky, August, 1978.

Filstead, W.J. Qualitative Methods: A Needed Perspective in Evaluation Research, In T.D. Cook, C.S. Reichardt, (Eds.), Qualitative and Quantitative Methods in Evaluation Research. Beverly Hills, CA: Sage, 1979.

Franchak, S.J., Franken, M.E., & Subisak, J. Specifications for Longitudinal Studies. Columbus, OH: The National Center for Research in Vocational Education, 1980.

References, continued

Franchak, S.J., & Spirer, J.E. Evaluation Handbook, Vol. 1: Guidelines and Practices for Follow-up Studies of Former Vocational Education Students. Columbus, OH: The National Center for Research in Vocational Education, 1978.

Grasso, J.T. Impact Evaluation: The State of the Art. Columbus, OH: The National Center for Research in Vocational Education, 1979.

Guba, E.G. Criteria for Assessing the Trustworthiness of Naturalistic Inquiries. In press.

Guba, E.G. Toward a Methodology of Naturalistic Inquiry in Educational Evaluation. Los Angeles, CA: Center for the Study of Evaluation, UCLA Graduate School of Education, University of California, 1978.

Guilford, J.P. Psychometric Methods. New York: McGraw Hill, 1954.

Horst, P., et al. Program management and the federal evaluator. Public Administration Review, July/August 1974, 300-308.

Kievit, M.B. Perspectivism in Choosing and Interpreting Outcome Measures in Vocational Education. Paper presented at the National Conference on Outcome Measures, Louisville, Kentucky, August, 1978.

References, continued

Lecht, L.A. Evaluating Vocational Education-Policies and Plans for the 1970s. New York: Praeger, 1974.

Lee, A.M. Use of Evaluative Data by Vocational Educators. Columbus, OH: The National Center for Research in Vocational Education, 1979.

McCaslin, N.L., Gross, C.J., & Walker, J.P. Career Education Measures: A Compendium of Evaluation Instruments. Columbus, OH: The National Center for Research in Vocational Education, 1979.

McCaslin, N.L., & Walker, J.P. A Guide for Improving Locally Developed Career Education Measures. Columbus, OH: The National Center for Research in Vocational Education, 1979.

Morell, J. Follow-up Research as an Evaluation Strategy: Theory and Methodologies. In T. Abramson, et al. (Eds.), Handbook of Vocational Education Evaluation. Beverly Hills, CA: Sage, 1979.

Nunnally, J.C. Psychometric Theory (2nd ed.). New York: McGraw Hill, 1978.

Patton, M.Q. Qualitative Evaluation Methods. Beverly Hills, CA: Sage, 1980.

References, continued

Pearsol, J.A. Combining Quantitative and Qualitative Educational Research Methods: A Perspective. Paper presented at the Annual Meeting of the American Vocational Association, New Orleans, Louisiana, December, 1980.

Pucel, D.J. Longitudinal Methods as Tools for Evaluating Vocational Education. Columbus, OH: The National Center for Research in Vocational Education, 1979.

Riffel, R. The Case Study and Its Use on the National Institute of Education Vocational Education Study. Paper presented at the Annual Meeting of the American Vocational Association, New Orleans, Louisiana, December, 1980.

Ruby, J. Exposing yourself: Reflexivity, anthropology, and film. Semiotica, 1980, 30, 153-79.

Scriven, M. Product Evaluation. Research and Evaluation Program Paper and Report Series No. 29. Portland, OR: The Program, Northwest Regional Educational Laboratory, 1979.

Simon, H.A. Administrative Behavior (3rd ed.). New York: The Free Press, 1976.

References, continued

Spirer, J.E. The Case Study Method: Guidelines, Practices and Applications for Vocational Education Evaluation. Columbus, OH: The National Center for Research in Vocational Education, 1980.

Steiner, E.S. Logical and Conceptual Analytic Techniques for Educational Researchers. Washington, DC: University Press of America, 1978.

Trend, M.G. On the Reconciliation of Qualitative and Quantitative Analyses: A Case Study. In T.D. Cook, C.S. Reichardt (Eds.), Qualitative and Quantitative Methods in Evaluation Research. Beverly Hills, CA: Sage, 1979.

Webb, E.J., et al. Unobtrusive Measures. Chicago: Rand McNally, 1966.

Wholey, J.S. Evaluability Assessment. In L. Rutman (Ed.), Evaluation Research Methods: A Basic Guide. Beverly Hills, CA: Sage, 1977.

Wholey, J.S. Evaluation: When is it really needed? Evaluation, 1975, 2, 89-93.